Statistically Learning Dispersed New Physics at the LHC

Sahana Narasimha

in collaboration with:
Mohammad Mahdi Altakach, Sabine Kraml, Andre Lessa, Timothée Pascal,
Humberto Reyes-Gonzalez, Wolfgang Waltenberger, Jamie Yellen

universität wien

HEPHY
INSTITUT FÜR HOCHENERGIEPHYSIK

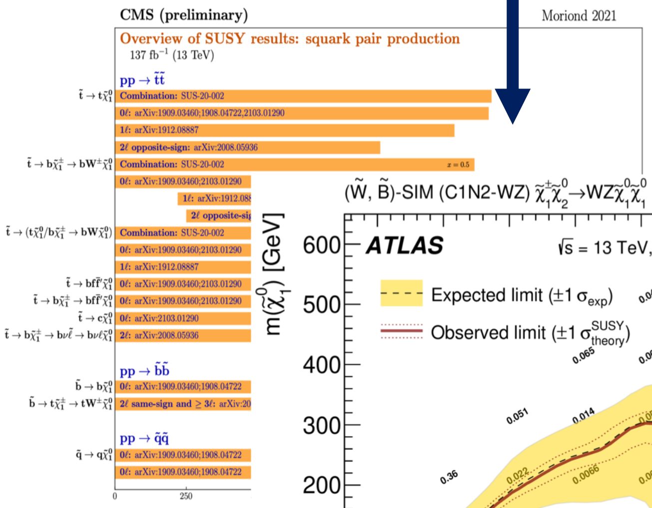
EUROPEAN PHYSICAL SOCIETY EPS HEP2025 MARSEILLE

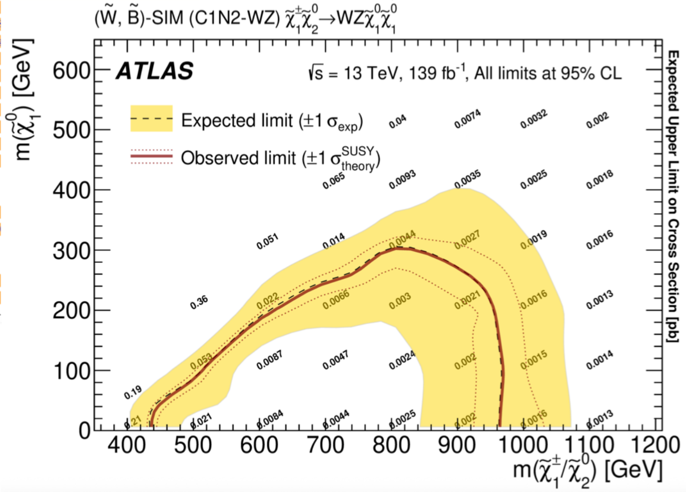
Motivation

A global view of what experimental data tells us about new physics

- LHC currently has no clear sign of where new physics might lie
- However, dispersed signals might be hiding in the slew of LHC data

Effects of new physics that are spread out over several search regions or final states



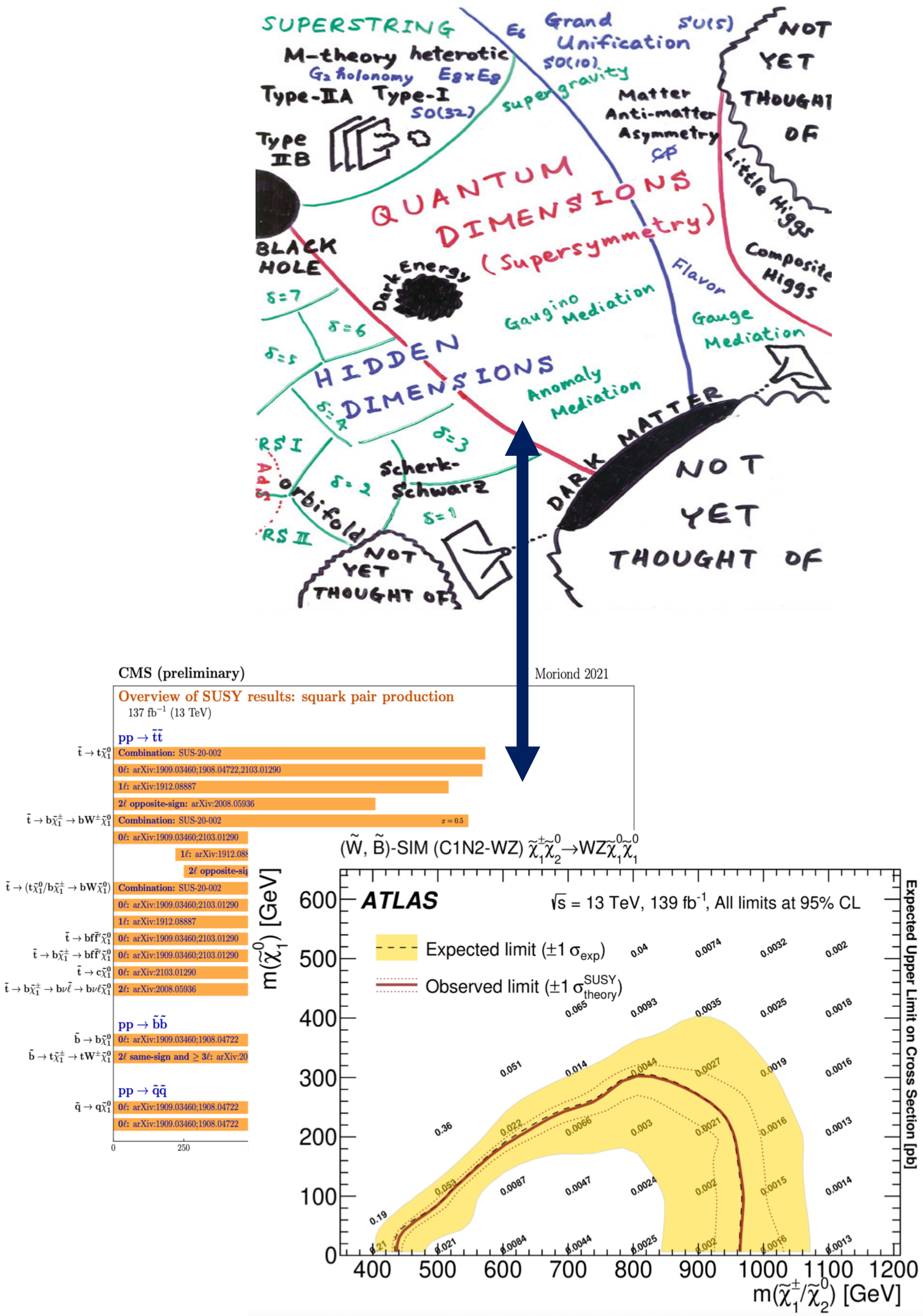


[1] Illustration from Hitoshi Murayama
[2] https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsSUS#Run_2_Summary_Plots_13_TeV, https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/SUSY-2018-41/figaux_07b.png

Motivation

A global view of what experimental data tells us about new physics

- LHC has currently no clear sign of where new physics may lie
- However, dispersed signals might be hiding in the slew of LHC data
- These are easily missed in the usual channel-by-channel analysis or disregarded as statistical fluctuations
- Change of perspective: a global exploration of LHC data to complement individual final state signature analyses



Top-Down Approach

Bottom-Up Approach

Top-Down Approach:

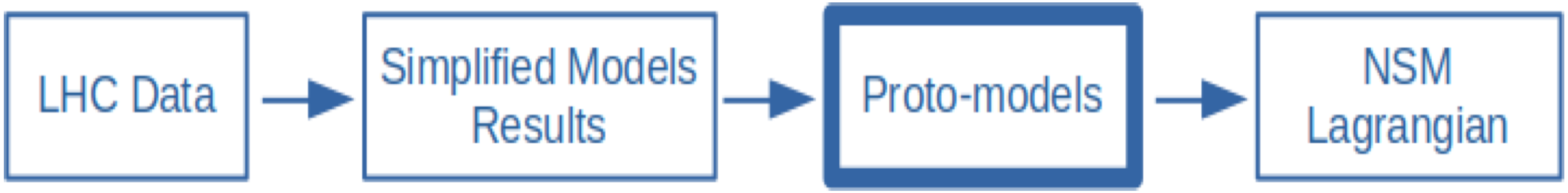
1. Construct a NP Lagrangian
2. Compute physical observables such as masses, cross sections,..
3. Compare with data -> get significance/ limits on observables

Bottom-Up Approach:

1. Take data
2. Fit (learn) physical observables - masses, cross sections, ….
3. Infer NP Lagrangian

Bottom-up Approach

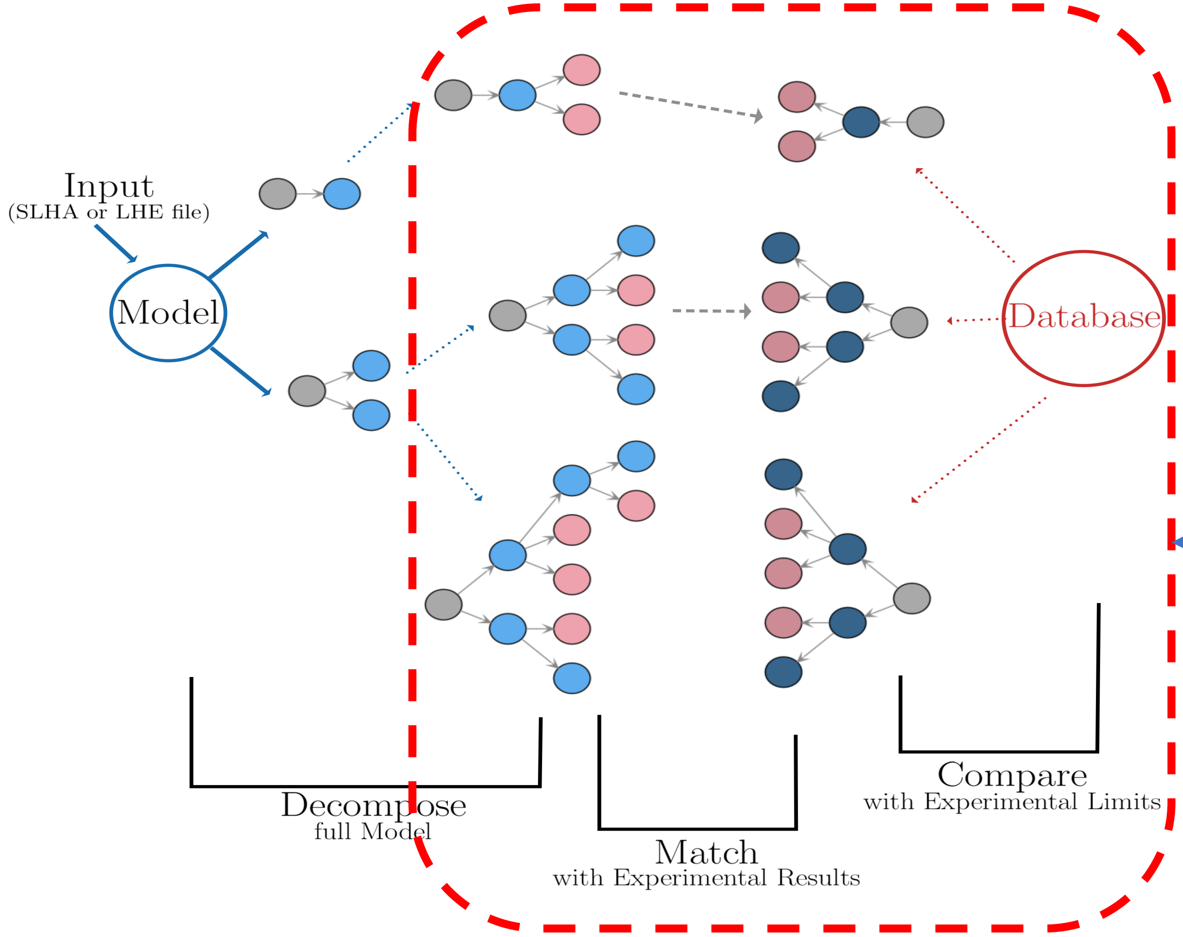
Given the data, can we build the next SM Lagrangian?



- Develop a statistical learning algorithm that identifies **potential excesses** amongst the published LHC data, while being compatible with all the constraints
- Build candidate **‘proto-models’** from them
- Based on simplified model results → exploit SModelS functionality and database

SModelS Working Principle

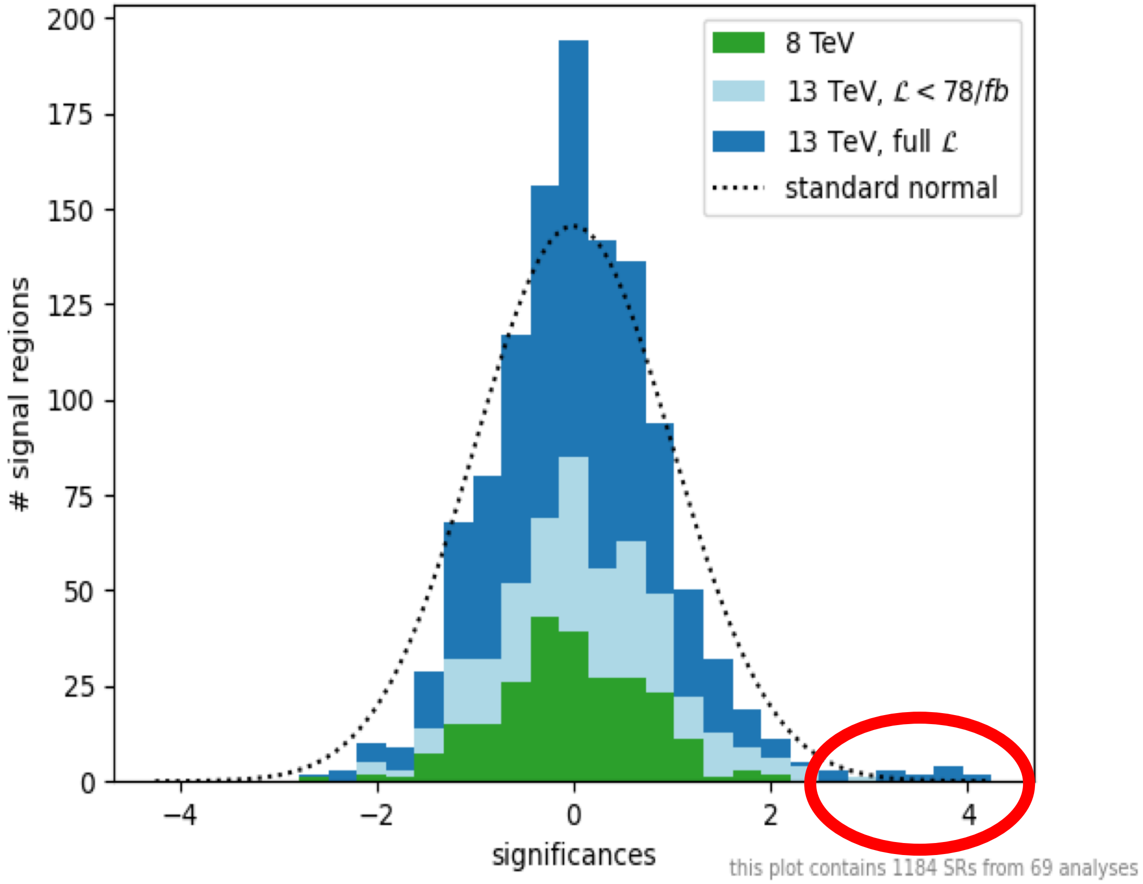
- Public tool which allows for a fast reinterpretation of LHC experimental results
- Confronts BSM signals against simplified model results from the LHC.



Exploit this part of the SModelS framework

Data

- Experimental constraints from around 110 published LHC results
- Database distribution of signal region significances under the SM hypothesis, representing observed deviations from SM predictions in units of the SM prediction uncertainty
- Notable deviation from the expected standard normal distribution, particularly in the right tail, indicating an excess of upward fluctuations beyond SM expectations



Proto-model…?

- Can be thought of as stacks/sets of simplified models - physics objects designed to capture experimental observations
- Not intended to be fully consistent theoretical models - properties are not bounded by higher level theoretical assumptions (such as gauge symmetries)
- Particle content motivated by database consisting of mostly SUSY-based simplified model searches
- Particle masses, production cross-sections and branching ratios treated as free parameters

Proto-model Particle Content	
Quark Partners Light - X_q $(q = u, d, c, s)$ Heavy- $X_t^{1,2}, X_b^{1,2}$	Gluon Partner X_g
Electroweak Partners $X_W^{1,2}, X_Z^{1,2,3}$ (LSP[1]= X_Z^1)	Lepton Partners X_l, X_{ν_l}

[1] Lightest Stable Particle

The algorithm

- A prototype of the proto-modelling approach was published in a proof-of-concept paper →
- The initial algorithm followed a MCMC-type random walk, freely adding/removing particles and other parameters in a varying dimensional state space
- We now extend the algorithm adapting ideas from **Reversible Jump Markov Chain Monte Carlo** algorithm, first proposed by Green[1]

Artificial Proto-Modelling: Building Precursors of a Next Standard Model from Simplified Model Results

Wolfgang Waltenberger,[a,b,1] André Lessa,[c] Sabine Kraml[d]

[a] *Institut für Hochenergiephysik, Österreichische Akademie der Wissenschaften, Nikolsdorfer Gasse 18, A-1050 Wien, Austria*

[b] *University of Vienna, Faculty of Physics, Boltzmanngasse 5, A-1090 Wien, Austria*

[c] *Centro de Ciências Naturais e Humanas, Universidade Federal do ABC, Santo André, SP - Brasil*

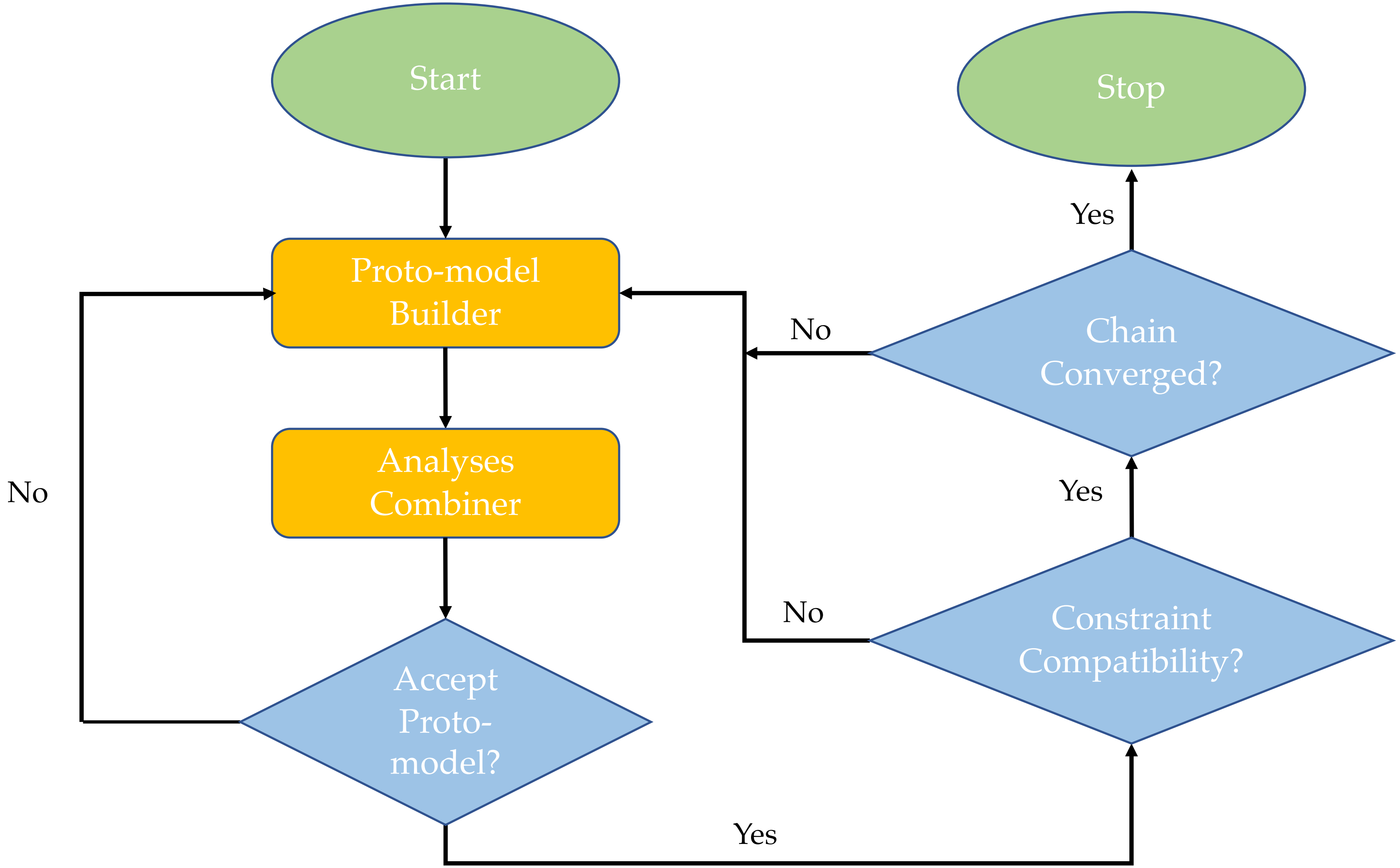
[d] *Laboratoire de Physique Subatomique et de Cosmologie, Université Grenoble-Alpes, CNRS/IN2P3, 53 Avenue des Martyrs, F-38026 Grenoble, France*

E-mail: wolfgang.waltenberger@oeaw.ac.at, andre.lessa@ufabc.edu.br, sabine.kraml@lpsc.in2p3.fr

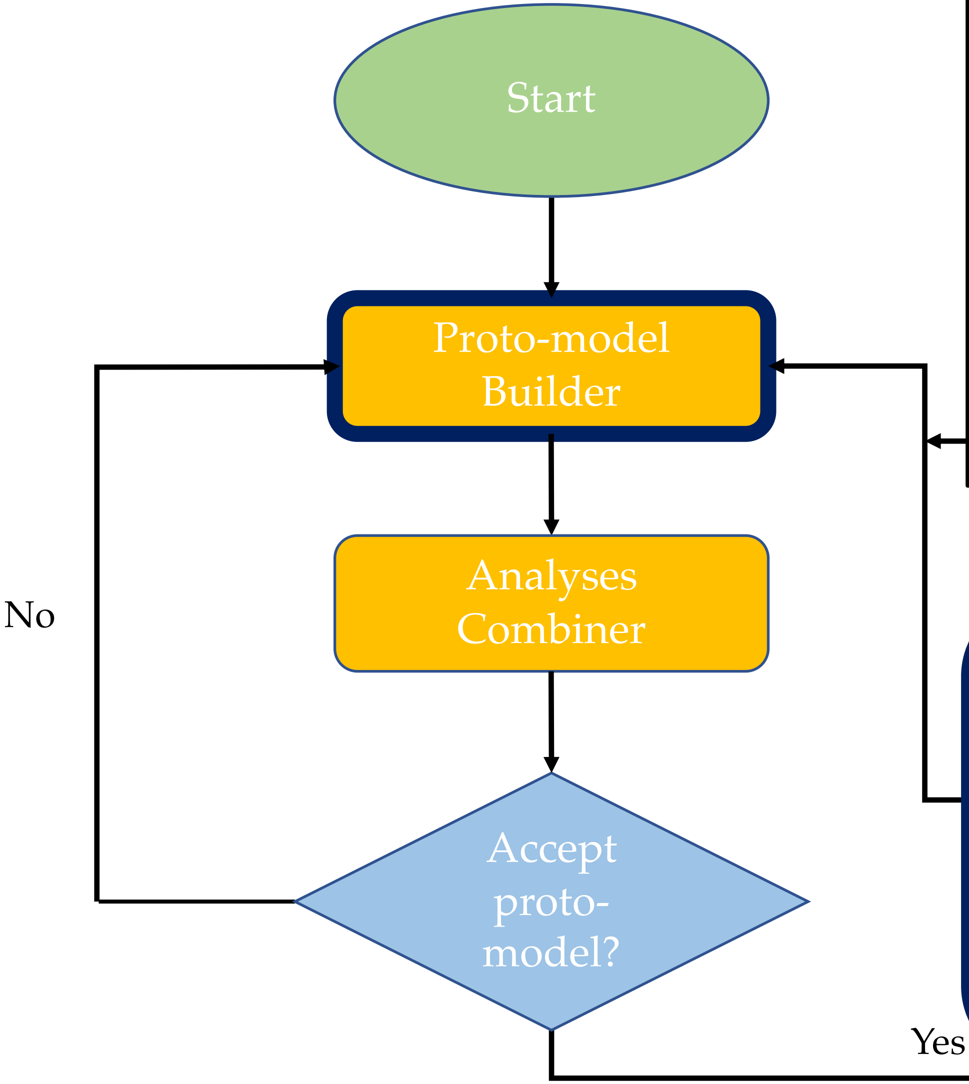
https://arxiv.org/pdf/2012.12246

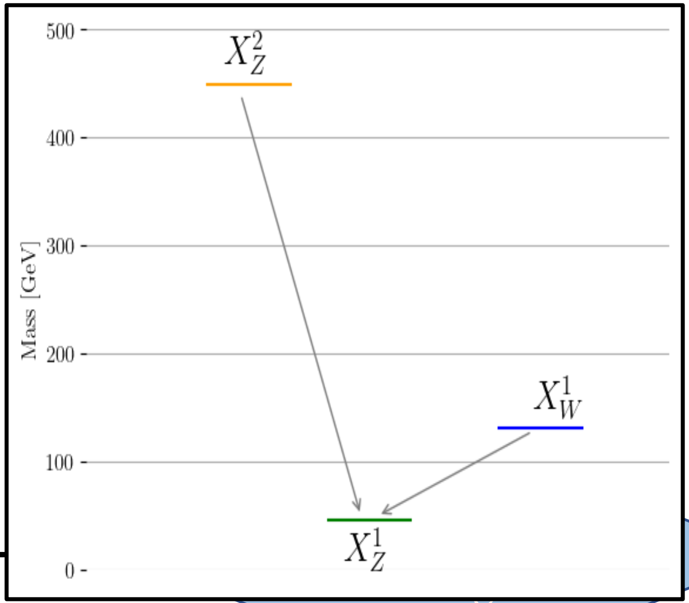
[1] Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, P. Green, https://doi.org/10.2307/2337340

Algorithm Flowchart



Algorithm Flowchart

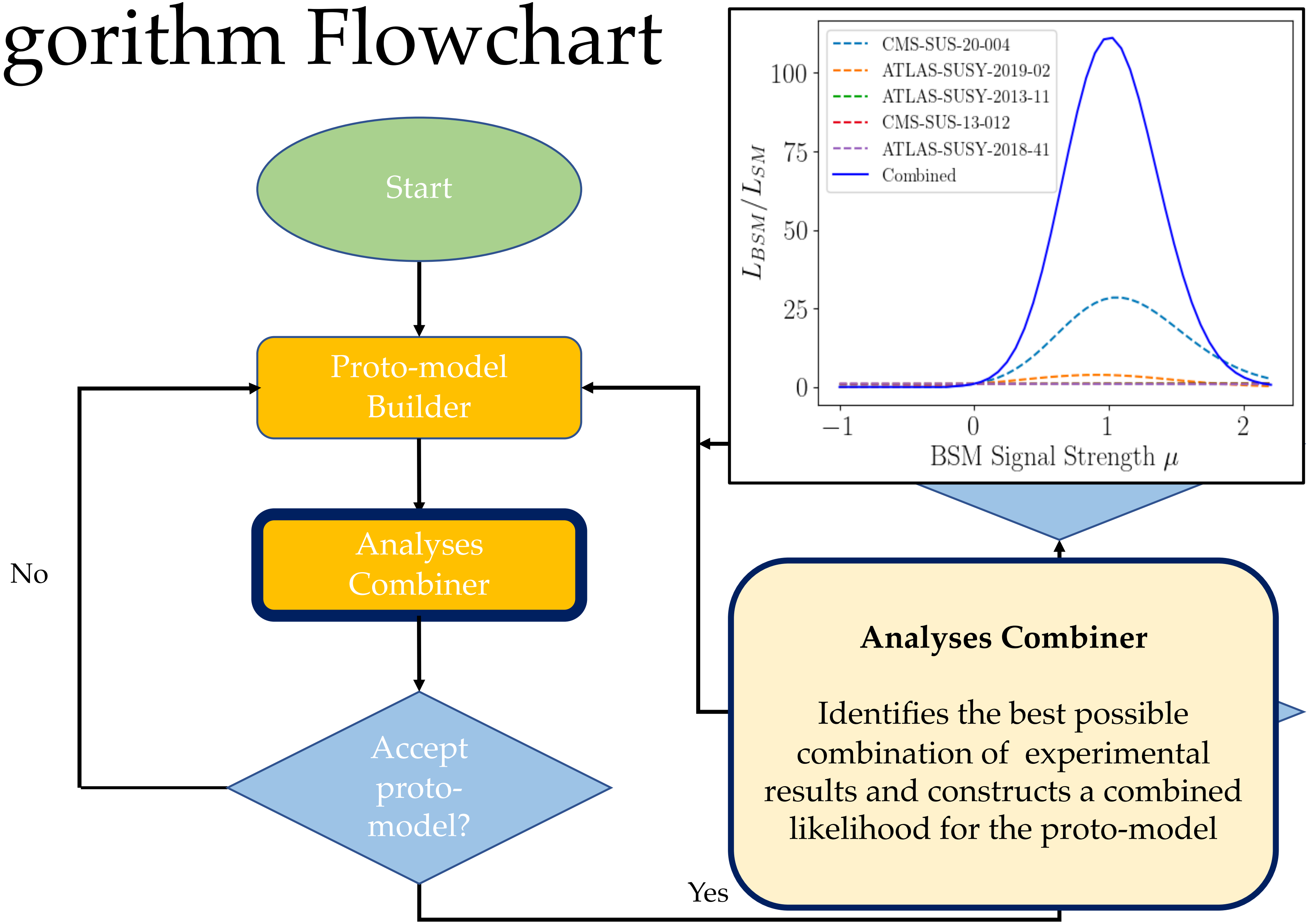




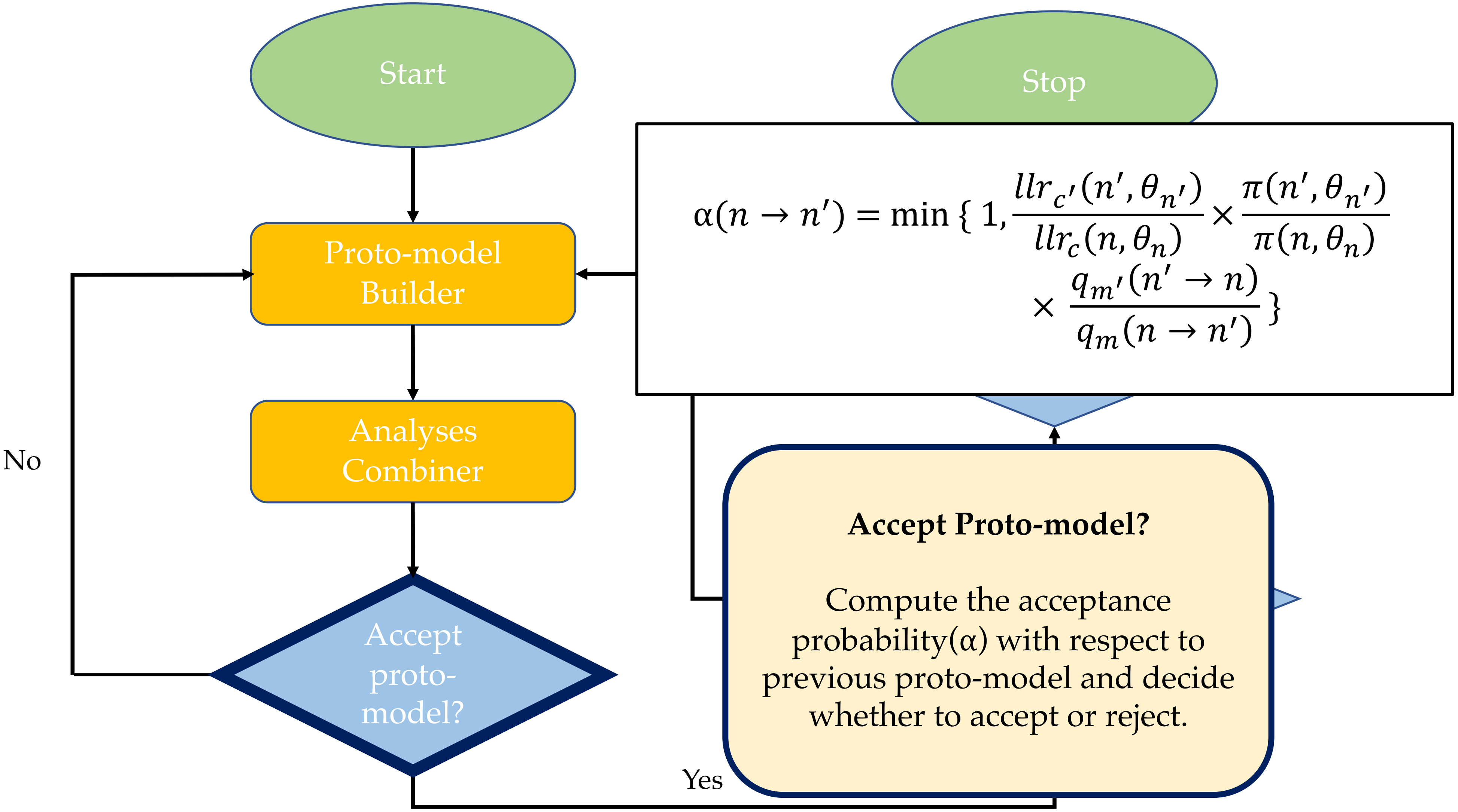
Proto-model Builder

Creates proto-models, randomly adding or removing new particles and changing any of the proto-model parameters.

Algorithm Flowchart



Algorithm Flowchart

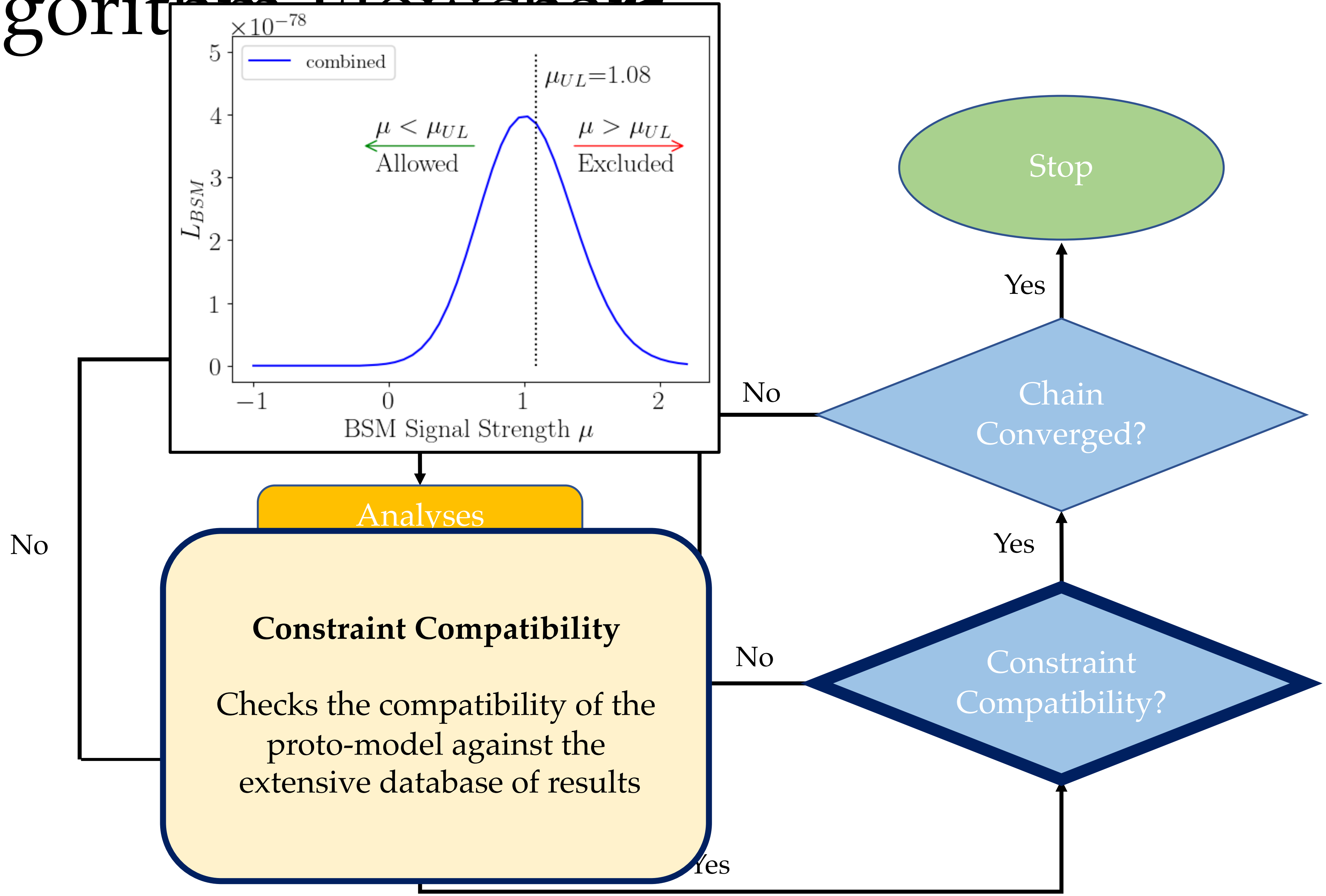


$$\alpha(n \to n') = \min\left\{ 1, \frac{llr_{c'}(n', \theta_{n'})}{llr_c(n, \theta_n)} \times \frac{\pi(n', \theta_{n'})}{\pi(n, \theta_n)} \times \frac{q_{m'}(n' \to n)}{q_m(n \to n')} \right\}$$

Accept Proto-model?

Compute the acceptance probability(α) with respect to previous proto-model and decide whether to accept or reject.

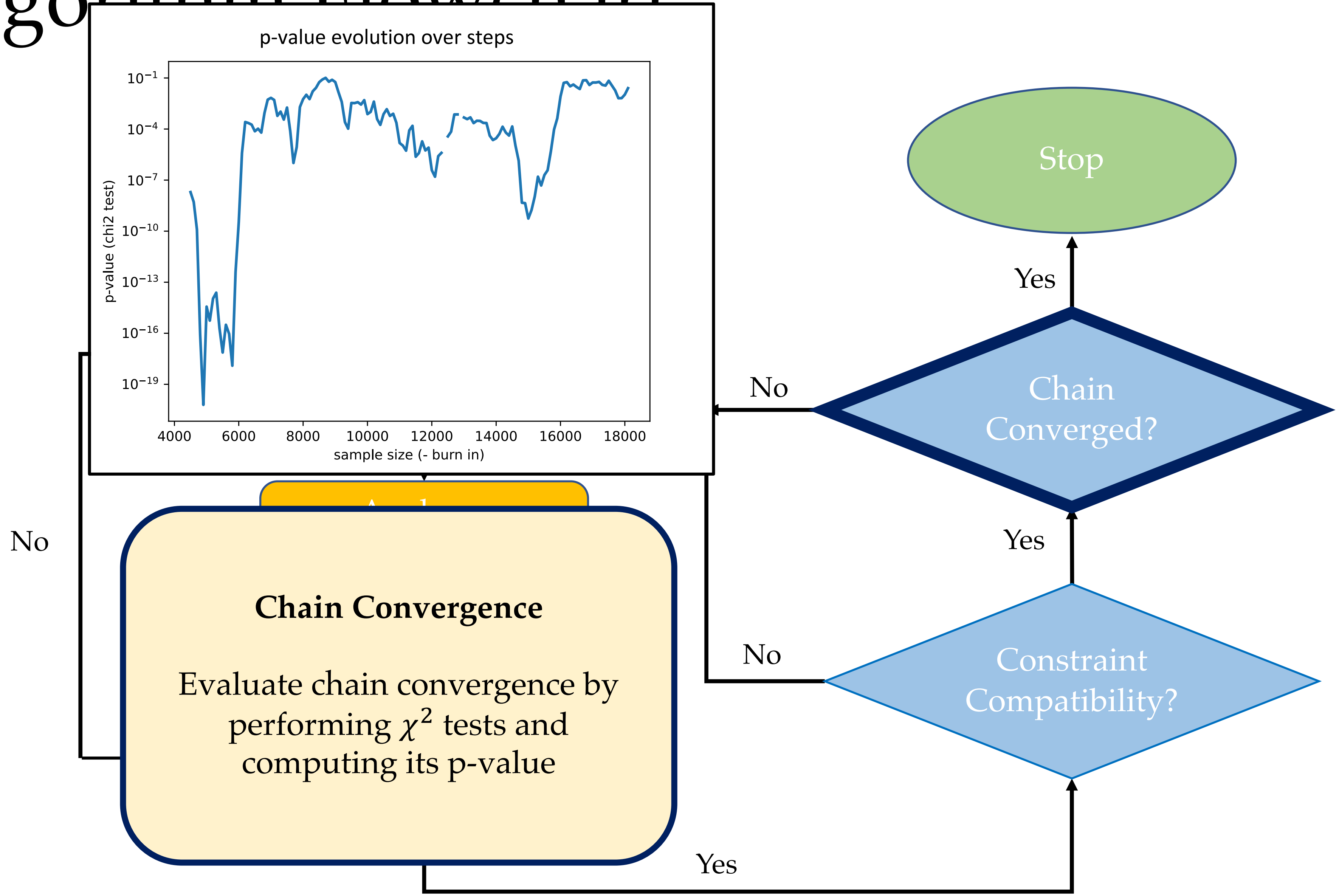
Algorithm Flowchart



Constraint Compatibility

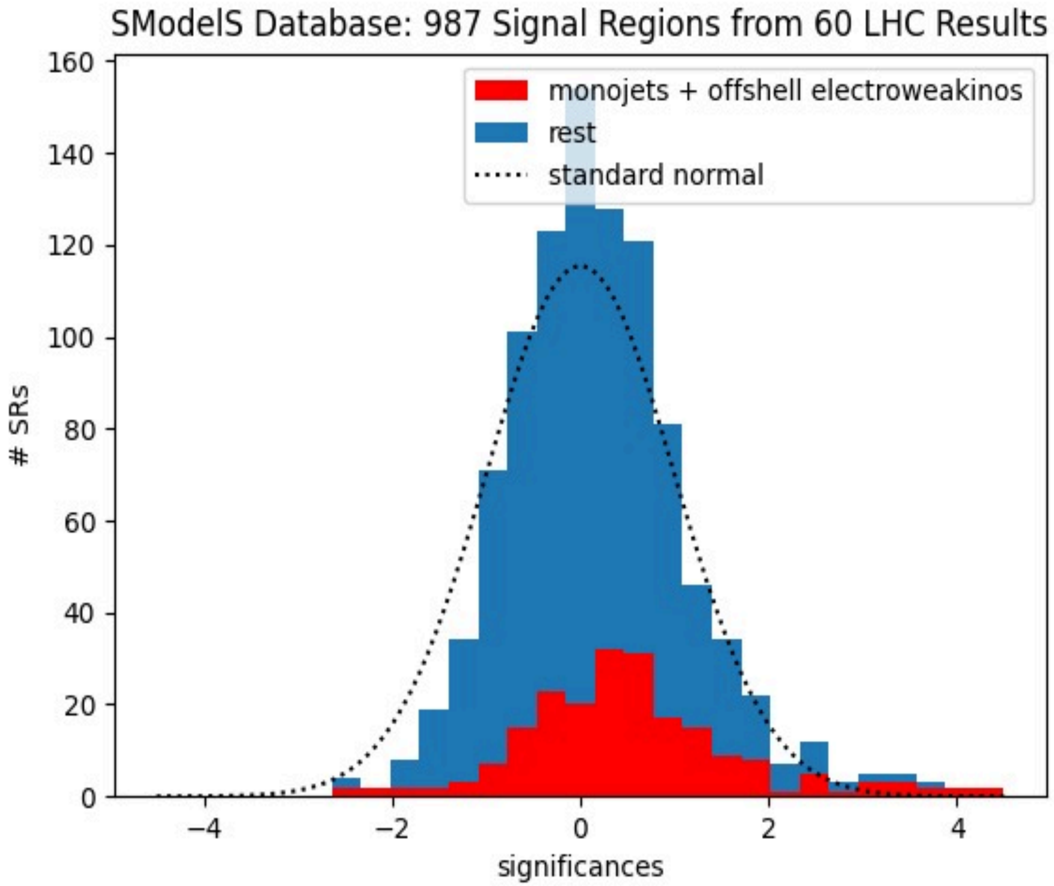
Checks the compatibility of the proto-model against the extensive database of results

Algorithm Flowchart



Run on SModelS Database

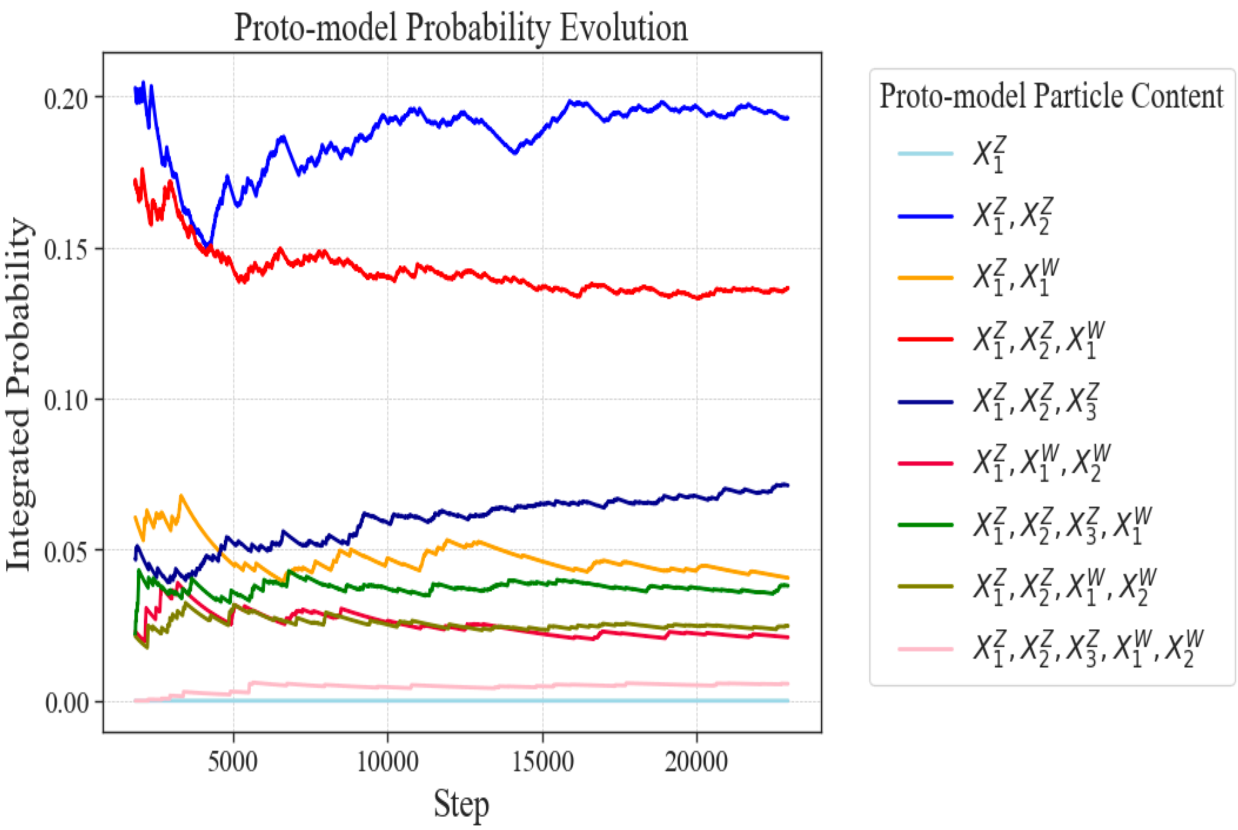
- Algorithm executed over a subset of the proto-model space – considering only electroweak partner particles
- This phase space produces signatures that exhibit more excesses compared to the full result set

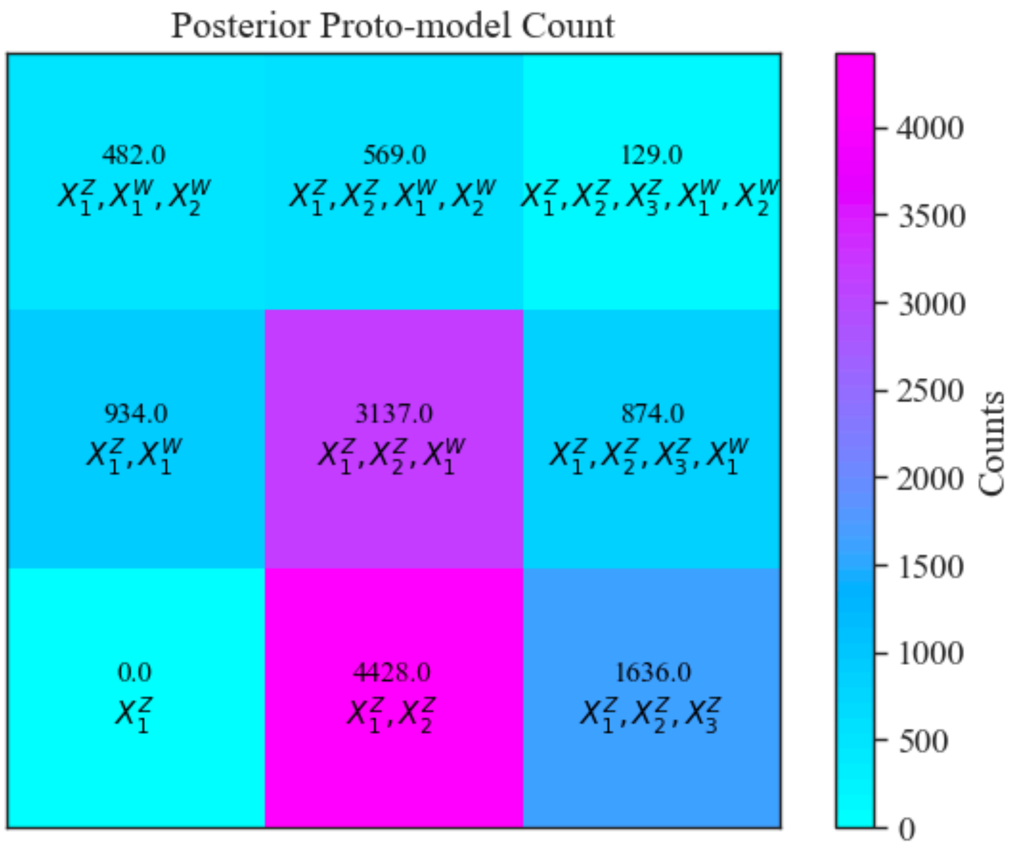


Preliminary Results

Posterior Phase Space

Probability of proto-models generated across a walk over the electroweak partner particles

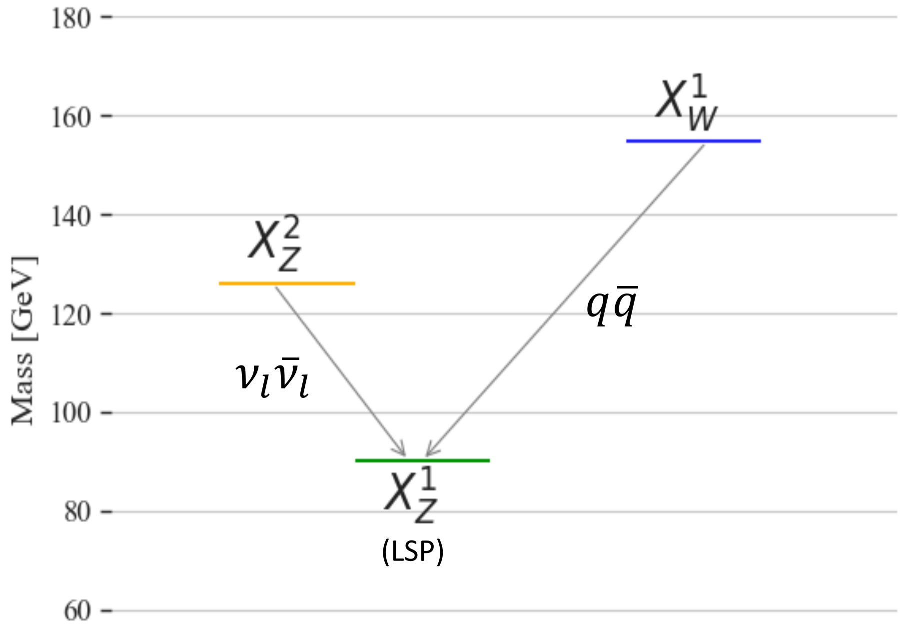




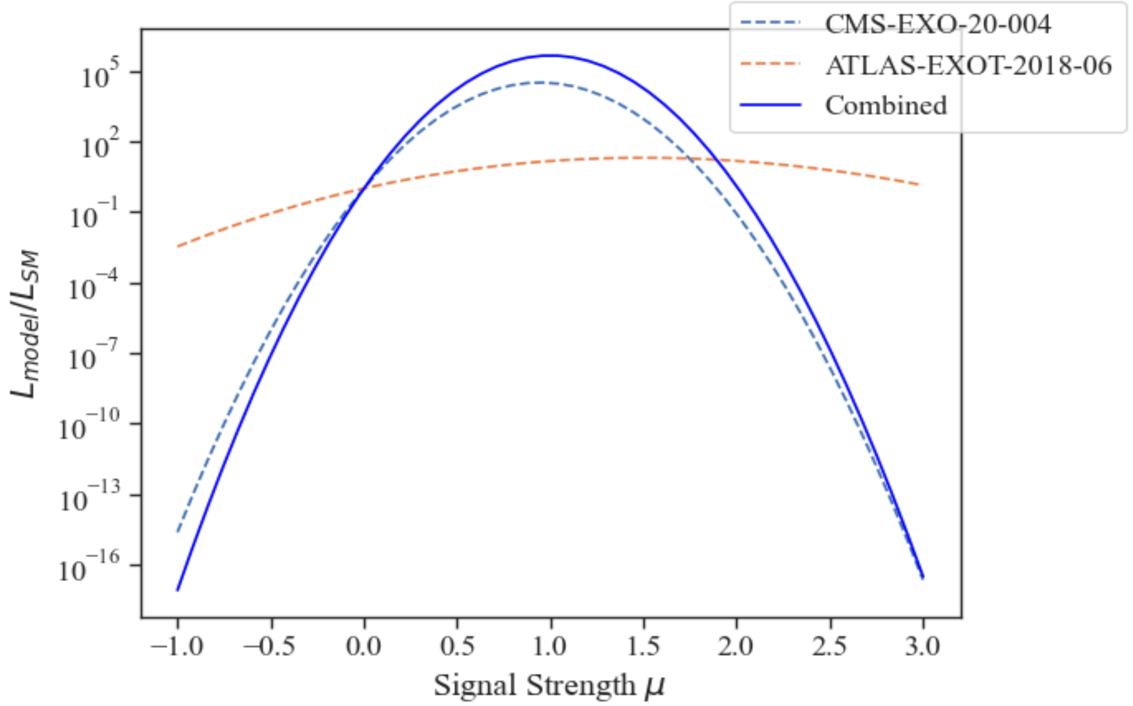
Current best proto-model

Current proto-model with the maximum deviation from the SM expectation

Production xsecs: (in units of corresponding SUSY xsecs):
$((X_Z^1, X_Z^2)$: 0.61, $(X_Z^1, X_W^{1\pm})$: 1.3, $(X_Z^2, X_W^{1\pm})$: 0.23, (X_W^{1-}, X_W^{1+}) : 2.12



Proto-model generation driven by two monojet searches – 'CMS-EXO-20-004' and 'ATLAS-EXOT-2018-16'



Summary

- A new approach to data driven search for new physics based on simplified model results, using SModelS framework to construct ”proto-models”.
- Proto-models → not fully consistent theoretical models, but try to capture excesses in data.
- Algorithm borrows ideas from RJMCMC methods, with additional entities – proto-model builder, analysis combiner and a check against experimental constraints.
- Preliminary runs over the electroweak partner particles favour proto-models with two neutral electroweak partners and/or one charged electroweak partner.
- Excesses in monojet searches drive proto-model generation with the maximum deviation from SM expectation.
- Closure tests of the algorithm and its results are in progress.

Thank You!

Back Up

The algorithm

- Follows a random walk, freely adding/removing particles and other parameters in a varying dimensional state space
- Adapt ideas from Reversible Jump Markov Chain Monte Carlo algorithm, first proposed by Green
- RJMCMC method retains the reversibility and detailed balance properties of the Metropolis-Hastings algorithm.
- Requires "dimension matching" while constructing the proposal densities to account for differences in dimension of different subspaces

Metropolis Hastings detailed balance condition in MCMC:

$$P(x)q(x \to y) = P(y)q(y \to x)$$

The algorithm

- For our case, we require the following detailed balanced condition to be upheld

$$llr_c(M_i)\times\pi(M_i)\times q(M_i \to M_j) = llr_{c'}(M_j)\times\pi(M_j)\times q(M_j \to M_i) \quad (3)$$

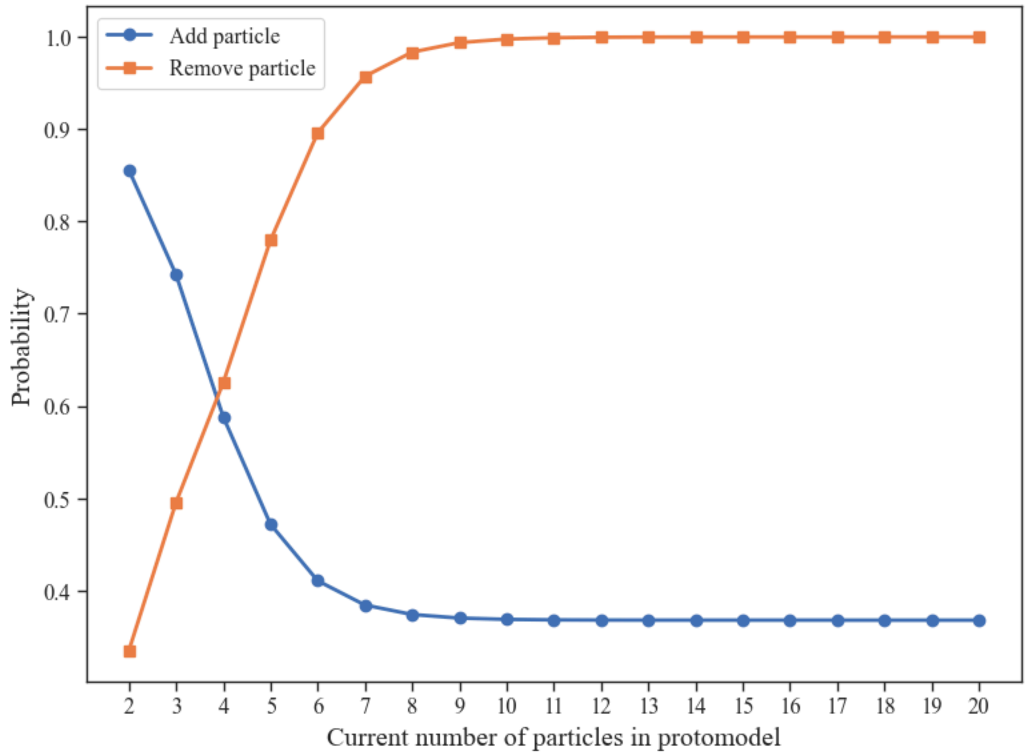
$llr_c = \frac{L_c(\mu=1|M)}{L_c(\mu=0|M)}$, where c refers to the combination of experimental results for which the likelihood is being built

$\pi(M)$ = Prior on the particular proto-model

q = Proposal density for jumping between different proto-model states

Proto-model Builder

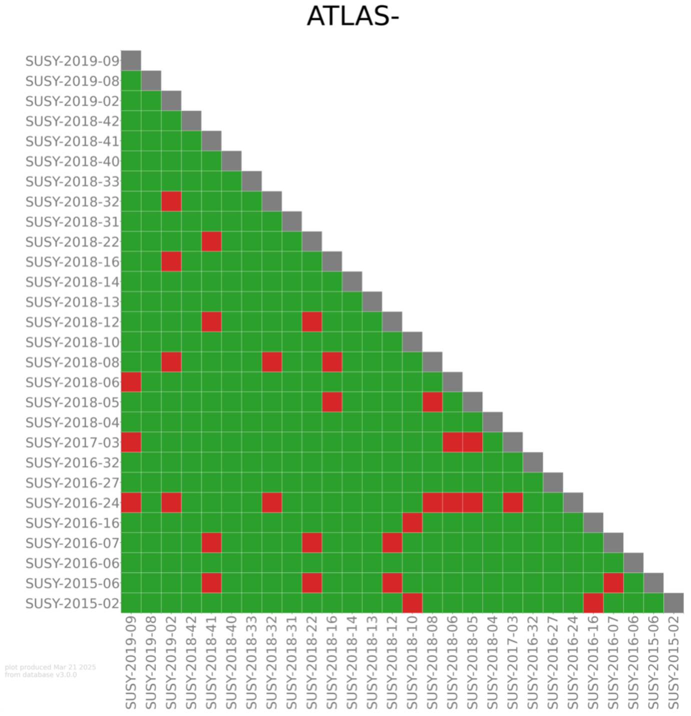
- The builder block is responsible for random changes in the proto-model.
- Here we attempt multiple move types and define the proposal densities for each move
    - Changing the dim of the proto-model – define (q_{add}, q_{rem})
        - We add/remove parameters to/from the proto-model with probabilities q_{add}/q_{rem} . We could either add a new particle with all its corresponding parameters, or add extra parameters to the current particle content in the proto-model.
    - Changing the values of existing parameters in the proto-model
        - Construct proposal probabilities for each parameter to be changed

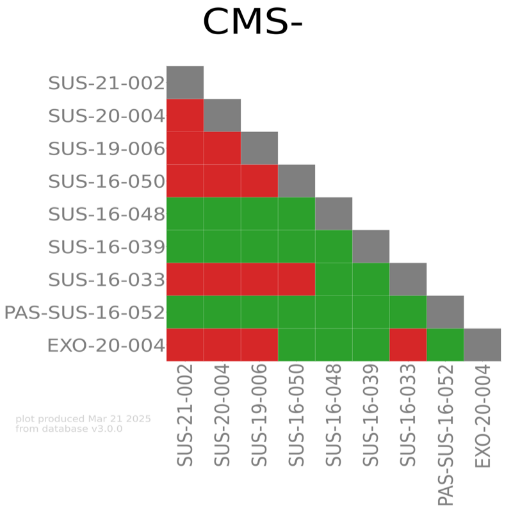


Analysis Combiner

- Initially build a combination matrix, that stores information of all the experimental results that can be considered to be approximately uncorrelated, and thus can be combined
- Collects all the experimental analyses which provides a result for the proto-model
- Use the “pathfinder”[1] algorithm, which takes in the set of theory predictions and the combination matrix and finds the combination with the most significant deviation from the SM, i.e

$$c' = max_{c \in C} \prod_i^c L_M^i(\mu = 1)/L_{SM}^i \qquad (3)$$





Test Statistic

To compare different proto-models with different degrees of freedom, we further define a test statistic K

$$K = 2\log\frac{L_M(\mu=1)\pi(M)}{L_{SM}\pi(SM)} \tag{4}$$

where π is a prior on the proto-model, that punishes the proto-model for unneeded complexity (principle of parsimony):

$$\pi(M) = \exp[-(\frac{n_{par}}{a_1} + \frac{n_{br}}{a_2} + \frac{n_{prod}}{a_3})], \qquad \pi(SM) = 1 \tag{5}$$

This test statistic roughly corresponds to a $\Delta\chi^2$ of the proto-model with respect to the SM, with a penalty for new degrees of freedom:

$$K \approx \Delta\chi^2 + 2\ \ln\pi(M) \tag{6}$$

Acceptance Probability

- To go from $M_i \to M_j$, we may attempt multiple move types indexed by m, the acceptance probability becomes:

$$\alpha(M_i \to M_j) = \min\left\{1, \underbrace{\frac{llr_{c'}(M_j)}{llr_c(M_i)}}_{\text{Likelihood Ratio}} \times \underbrace{\frac{\pi(M_j)}{\pi(M_i)}}_{\text{Prior Ratio}} \times \underbrace{\frac{q_{m'}(M_j \to M_i)}{q_m(M_i \to M_j)}}_{\text{Proposal Ratio}}\right\} \quad (7)$$

Likelihood Ratio × Prior Ratio: $\exp[\frac{1}{2}(K_j - K_i)]$

Likelihood Ratio = Ratios of $llr_c \equiv \frac{L_c(\mu=1|n,\theta_n)}{L_c(\mu=0|n,\theta_n)}$, where c refers to the combination of experimental results for which the combined likelihood is computed

Prior ratio = Ratios of prior probabilities on the proto-model M

Proposal ratio = Ratios of proposal densities to move from $M_{i/j} \to M_{j/i}$

Critic

- To get constraints on the input protomodel, SModelS computes an observable called "r-value"

$$r_{obs} = \sigma_M/\sigma_{UL}$$

- If $r_{obs} > 1$, we conclude the model point to be excluded

Fast Critic

- Quickly get upper-limit constraints on the proto-model
- Allow for a 30% violation of the upper limit, i.e if $r_{obs} > 1.3$, reject the proto-model
- Else compute the slow critic

Slow Critic

- More robust, but computationally more expensive
- Find the combined set of analyses which give the most sensitive constraints i.e the combination that minimizes

$$c' = min_{c \in C} \prod_i L_M^{i\,(exp)}(\mu = 1)/L_{SM}^{i\,(exp)}$$

- Compute r_{obs} of the combined likelihood
- If $r_{obs} > 1$, reject the proto-model

The Posterior phase space

Closure Tests (Yet to do..)

- Walk on fake-SM data, get distribution of test statistic K under the SM-only hypothesis, and compute global p-value for K
- Replace the observed data with the expected background plus signal yields for a given proto-model, run and see if the algorithm discovers this signal
- Proper convergence over the proto-model parameter space